REMARKS

Claims 1-17 are pending. Claims 1-4, 7-10, and 13-15 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,078,953 to Vaid. Claims 5-6, 11-12, and 16-20 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,078,953 to Vaid in view of U.S. Patent No. 5,276,677 to Ramamurthy.

Reconsideration is requested. No new matter is added. The specification is amended. Claims 1, 7, and 13 are amended. Claims 6, 12, and 17-20 are canceled. Claims 21-26 are added. The rejections are traversed. Claims 1-5, 7-11, 13-16, and 21-26 remain in the case for consideration.

The Examiner requested that the Applicant supply copies of the pertinent pages from the priority application Serial Number 09/512,963. Pages 4-12, 15-18 and FIGS. 4-5G are hereby attached for the Examiner's reference.

INTERVIEW SUMMARY

On April 21, 2004, the undersigned held a telephone interview with Examiner Lezak. The undersigned thanks the Examiner for taking the time to discuss the case and suggesting amendments that would aid in overcoming the rejections. Claims 1, 7, and 13 were discussed. Although no agreement was reached regarding the allowability of the claims as currently presented, the Examiner agreed that including in the claims the concept of the topological vector space as the source of the vectors in the template would help overcome the rejection over Vaid.

The Examiner also suggested including the concept of the "bounding circle" as shown in FIG. 2. The Examiner suggested that by making clear that the threshold distance defined a multidimensional area including vectors not in the template would distinguish the claims over the prior art. Unfortunately, the Applicant cannot make such an amendment. As stated in the specification at page 4, line 18, circle 210 of FIG. 2 is an abstraction, useable only if the template could be reduced to a single point in the multidimensional vector space. Since the template typically includes many vectors, such a reduction is not appropriate. It might happen that there is a set of vectors, all within "circle" 210, that would be more than the threshold distance from the template, and so including in the claims the concept of a "circle" would be a useless concept.

But the underlying concept suggested by the Examiner, that there are other sets of vectors within the threshold distance of the template, is a sound one, and is supported by the specification. Accordingly, claims 1, 7, and 13 have been amended to introduce the concept

of the impact summary as a subset of vectors from the topological vector space, which might not be coincidental to the vectors in the template. While this amendment is not the specific wording suggested by the Examiner, it is consistent with the Examiner's suggestion. The Applicant hopes that the Examiner will read the claims in the spirit presented.

The Examiner indicated that she believed other prior art existed that would either anticipate or make obvious the claimed invention. The Examiner requested that the Applicant conduct a search for additional prior art she believes exists. Although the Applicant does not believe other material prior art exists, the undersigned performed the requested search, searching the U.S. Patent & Trademark Office database using the terms "Microsoft" and "Hausdorff." The pertinent prior art found is hereby submitted on the accompanying Form PTO-1449. In the opinion of the Applicant, the prior art found by this search is no more material than the prior art already of record, and that the even in combination with the newly-submitted prior art, the claimed invention is neither anticipated nor obvious.

REJECTION OF CLAIMS UNDER 35 U.S.C. § 103(a)

As described above, in the telephone interview of April 21, 2004, the Examiner agreed to certain amendments that would overcome the prior art. Such amendments have been presented. Accordingly, the claims should now be allowable over the prior art of record.

For the foregoing reasons, reconsideration and allowance of claims 1-5, 7-11, 13-16, and 21-26 of the application as amended is solicited. The Examiner is encouraged to telephone the undersigned at (503) 222-3613 if it appears that an interview would be helpful in advancing the case.

Respectfully submitted,

MARGER JOHNSON & McCOLLOM, P.C.

Ariel S. Rogson

Ariel S. Rogson, Reg. No. 43,054

I HEREBY CERTIFY THAT THIS CORRESPONDENCE IS BEING DEPOSITED WITH THE UNITED STATES POSTAL SERVICE AS FIRST CLASS MAIL IN AN ENVELOPE ADDRESSED TO:

☐ COMMISSIONER OF PATENTS AND TRADEMARKS WASHINGTON D.C. 20231

☒ MAIL STOP RCE COMMISSIONER FOR PATENTS BOX 1450 ALEXANDRIA, VA 22313-1450

☐ BOX ______ COMMISSIONER FOR TRADEMARKS 2900 CRYSTAL DRIVE ARLINGTON, VA 22202-3513

ON: 16 June 2004

Ariel S. Rogson

MARGER JOHNSON & McCOLLOM, P.C.
1030 SW Morrison Street
Portland, OR 97205
503-222-3613
Customer No. 20575

to all other concepts identified by concept identification unit 130. Basis unit 140 is responsible for selecting a subset of the chains to form a basis for the directed set. Because basis unit 140 selects a subset of the chains established by chain unit 135, basis unit 140 is depicted as being part of chain unit 135. However, a person skilled in the art will recognize that basis unit 140 can be separate from chain unit 135. Measurement unit 145 is responsible for measuring how concretely each chain in the basis represents each concept. (How this measurement is performed is discussed below.) In the preferred embodiment, concept identification unit 130, chain unit 135, basis unit 140, and measurement unit 145 are implemented in software. However, a person skilled in the art will recognize that other implementations are possible. Finally, computer system 105 includes a data structure 150 (discussed with reference to FIG. 13 below). The data structure is responsible for storing the concepts, chains, and measurements of the directed set.

FIG. 1B shows computer system 105 connected over a network connection 140 to a network 145. The specifics of network connection 140 are not important, so long as the invention has access to a content stream to listen for concepts and their relationships. Similarly, computer system 105 does not have to be connected to a network 145, provided some content stream is available.

FIG. 2 shows computer system 105 listening to a content stream. In FIG. 2, network connection 140 includes a listening device 205. Listening device 205 (sometimes called a "listening mechanism") allows computer system 105 to listen to the content stream 210 (in FIG. 2, represented as passing through a "pipe" 215). Computer system 105 is parsing a number of concepts, such as "behavior," "female," "cat," "Venus Flytrap," "iguana," and so on. Listening device 205 also allows computer system 105 to determine the relationships between concepts.

But how is a computer, such as computer system 105 in FIGs. 1A, 1B, and 2 supposed to understand what the data it hears means? This is the question addressed below.

Semantic Value

Whether the data expressing content on the network is encoded as text, binary code, bit map or in any other form, there is a vocabulary that is either explicitly (such as for code) or implicitly (as for bitmaps) associated with the form. The vocabulary is more than an

arbitrarily-ordered list: an element of a vocabulary stands in relation to other elements, and the "place" of its standing is the *semantic value* of the element. For example, consider a spoon. Comparing the spoon with something taken from another scene – say, a shovel – one might classify the two items as being somewhat similar. And to the extent that form follows function in both nature and human artifice, this is correct! The results would be similar if the spoon were compared with a ladle. All three visual elements – the spoon, the shovel and the ladle – are *topologically* equivalent; each element can be transformed into the other two elements with relatively little geometric distortion.

What happens when the spoon is compared with a fork? Curiously enough, both the spoon and the fork are topologically equivalent. But comparing the ratio of boundary to surface area reveals a distinct contrast. In fact, the attribute (boundary)/(surface area) is a crude analog of the *fractal dimension* of the element boundary.

Iconic Representation

Fractal dimension possesses a nice linear ordering. For example, a space-filling boundary such as a convoluted coastline (or a fork!) would have a higher fractal dimension than, say, the boundary of a circle. Can the topology of an element be characterized in the same way? In fact, one *can* assign a topological measure to the vocabulary elements, but the measure may involve aspects of homotopy and homology that preclude a simple linear ordering. Suppose, for visual simplicity, that there is some simple, linearly ordered way of measuring the topological essence of an element. One can formally represent an *attribute space* for the elements, where *fork-like* and *spoon-like* resolve to different regions in the attribute space. In this case, one might adopt the standard Euclidean metric for $\mathbb{R}^2$ with one axis for "fractal dimension" and another for "topological measure," and thus have a well-defined notion of *distance* in attribute space. Of course, one must buy into all the hidden assumptions of the model. For example, is the orthogonality of the two attributes justified, i.e., are the attributes truly independent?

The example attribute space is a (simplistic) illustration of a semantic space, also known as a *concept space*. Above, the concern was with a vocabulary for human visual elements: a kind of *visual lexicon*. In fact, many researchers have argued for an *iconic representation* of meaning, particularly those looking for a representation unifying perception

and language. They take an empirical positivist position that *meaning* is simply an artifact of the "binding" of language to perception, and point out that all writing originated with pictographs (even the letter "A" is just an inverted ox head!). With the exception of some very specialized vocabularies, it is an unfortunate fact that most iconic models have fallen well short of the mark. What is the visual imagery for the word "maybe"? For that matter, the above example iconic model has shown how spoons and forks are different, but how does it show them to be the same (i.e., cutlery)?

Propositional Representation

Among computational linguists, a leading competitive theory to iconic representation is *propositional representation*. A proposition is typically framed as a pairing of an *argument* and a *predicate*. For example, the fragment "a red car" could be represented propositionally as the argument "a car" paired with the predicate "is red." The proposition simply asserts a property (the predicate) of an object (the argument). In this example, stipulating the argument alone has consequences; "a car" invokes the existential quantifier, and asserts instances for all relevant primitive attributes associated with the lexical element "car."

How about a phrase such as "every red car"? Taken by itself, the phrase asserts nothing – not even existence! It is a null proposition, and can be safely ignored. What about "every red car has a radio"? This is indeed making an assertion of sorts, but it is asserting a property of the semantic space itself; i.e., it is a meta-proposition. One can not instantiate a red car without a radio, nor can one remove a radio from a red car without either changing the color or losing the "car-ness" of the object. Propositions that are interpreted as assertions rather than as descriptions are called "meaning postulates."

At this point the reader should begin to suspect the preeminent role of the predicate, and indeed would be right to do so. Consider the phrase, "the boy hit the baseball."

nominative:	the boy	→ (is human), (is ~adult), (is male), (is ~infant), etc.
predicate:	(hit the baseball)	→
verb:	hit	→ (is contact), (is forceful), (is aggressive), etc.
d.o.:	the baseball	→ (is round), (is leather), (is stitched), etc.

The phrase has been transformed into two sets of attributes: the nominative attributes and two subsets of predicate attributes (verb and object). This suggests stipulating that all propositions must have the form (n: $n \in N$, p: $p \in P$), where N (the set of nominatives) is some appropriately restricted subset of $\wp$(P) (the power set of the space P of predicates). N is restricted to avoid things like ((is adult) and (is ~adult)). In this way the predicates can be used to generate a semantic space. A semantic representation might even be possible for something like, "The movie *The Boy Hit the Baseball* hit this critic's heart-strings!"

Given that propositions can be resolved to sets of predicates, the way forward becomes clearer. If one were to characterize sets of predicates as *clusters* of points in an attribute space along with some notion of *distance* between clusters, *one could quantify how close any two propositions are to each other.* This is the Holy Grail.

Before leaving this section, observe that another useful feature of the propositional model is *hierarchy of scope*, at least at the sentence level and below. Consider the phrase, "the boy hit the spinning baseball." The first-tier proposition is "x hit y." The second-tier propositions are "x is-a boy," and "y is-a baseball." The third-tier proposition is "y is spinning." By restricting the scope of the semantic space, attention can be focused on "hitting," "hitting spinning things," "people hitting things," etc.

Hyponymy & Meaning Postulates - Mechanisms for Abstraction

Two elements of the lexicon are related by *hyponymy* if the meaning of one is included in the meaning of the other. For example, the words "cat" and "animal" are related by hyponymy. A cat *is* an animal, and so "cat" is a hyponym of "animal."

A particular lexicon may not explicitly recognize some hyponymies. For example, the words "hit," "touch," "brush, " "stroke, " "strike," and "ram" are all hyponyms of the *concept* "co-incident in some space or context." Such a concept can be formulated as a *meaning postulate*, and the lexicon is extended with the meaning postulate in order to capture formally the hyponymy.

Note that the words "hit" and "strike" are also hyponyms of the word "realize" in the popular vernacular. Thus, lexical elements can surface in different hyponymies depending on the inclusion chain that is followed.

Topological Considerations

Now consider the *metrization problem*: how is the distance between two propositions determined? Many people begin by identifying a set S to work with (in this case, S = P, the set of predicates), and define a *topology* on S. A topology is a set O of subsets of S that satisfies the following criteria:

- Any union of elements of O is in O.
- Any finite intersection of elements of O is in O.
- S and the empty set are both in O.

The elements of O are called the *open* sets of S. If X is a subset of S, and *p* is an element of S, then *p* is called a *limit point* of X if every open set that contains *p* also contains a point in X distinct from *p*.

Another way to characterize a topology is to identify a *basis* for the topology. A set B of subsets of S is a basis if

- S = the union of all elements of B,
- for $p \in b_\alpha \cap b_\gamma$, ($b_\alpha, b_\gamma \in B$), there exists $b_\lambda \in B$ such that $p \in b_\lambda$ and $b_\lambda \supseteq b_\alpha \cap b_\gamma$.

A subset of S is open if it is the union of elements of B. This defines a topology on S. Note that it is usually easier to characterize a basis for a topology rather than to explicitly identify all open sets. The space S is said to be *completely separable* if it has a countable basis.

It is entirely possible that there are two or more characterizations that yield the same topology. Likewise, one can choose two seemingly closely-related bases that yield nonequivalent topologies. As the keeper of the Holy Grail said to Indiana Jones, "Choose wisely!"

The goal is to choose as *strong* a topology as possible. Ideally, one looks for a compact metric space. One looks to satisfy separability conditions such that the space S is guaranteed to be *homeomorphic* to a subspace of Hilbert space (i.e., there is a continuous and one-to-one mapping from S to the subspace of Hilbert space). One can then adopt the Hilbert space metric. Failing this, as much structure as possible is imposed. To this end, consider the following axioms (the so-called "trennungaxioms").

- T_0. Given two points of a topological space S, at least one of them is contained in an open set not containing the other.

- T_1. Given two points of S, each of them lies in an open set not containing the other.
- T_2. Given two points of S, there are *disjoint* open sets, each containing just one of the two points (Hausdorff axiom).
- T_3. If C is a closed set in the space S, and if p is a point not in C, then there are disjoint open sets in S, one containing C and one containing p.
- T_4. If H and K are disjoint closed sets in the space S, then there are disjoint open sets in S, one containing H and one containing K.

Note that a set X in S is said to be *closed* if the complement of X is open. Since the intention is not to take the reader through the equivalent of a course in topology, simply observe that the distinctive attributes of T_3 and T_4 spaces are important enough to merit a place in the mathematical lexicon – T_3 spaces are called regular spaces, and T_4 spaces are called normal spaces – and the following very beautiful theorem:

- **Theorem 1.** Every completely separable regular space can be imbedded in a Hilbert coordinate space.

So, if there is a countable basis for S that satisfies T_3, then S is metrizable. The metrized spaced S is denoted as (S, *d*).

Finally, consider $\mathcal{H}(\mathbf{S})$, the set of all compact (non-empty) subsets of $(\mathbf{S}, d)$. Note that for $u, v \in \mathcal{H}(\mathbf{S})$, $u \cup v \in \mathcal{H}(\mathbf{S})$; i.e., the union of two compact sets is itself compact. Define the pseudo-distance $\xi(x, u)$ between the point $x \in \mathbf{S}$ and the set $u \in \mathcal{H}(\mathbf{S})$ as

$$\xi(x, u) = \min\{d(x, y) : y \in u\}.$$

Using ξ define another pseudo-distance $\lambda(u, v)$ from the set $u \in \mathcal{H}(\mathbf{S})$ to the set $v \in \mathcal{H}(\mathbf{S})$:

$$\lambda(u, v) = \max\{\xi(x, v) : x \in u\}.$$

Note that in general it is *not* true that $\lambda(u, v) = \lambda(v, u)$. Finally, define the distance $h(u, v)$ between the two sets $u, v \in \mathcal{H}(\mathbf{S})$ as

$$h(u, v) = \max\{\lambda(u, v), \lambda(v, u)\}.$$

The distance function h is called the *Hausdorff* distance. Since

$$h(u, v) = h(v, u),$$

$$0 < h(u, v) < \infty \text{ for all } u, v \in \mathcal{H}(\mathbf{S}), u \neq v,$$

$$h(u, u) = 0 \text{ for all } u \in \mathcal{H}(\mathbf{S}),$$

$$h(u, v) \leq h(u, w) + h(w, v) \text{ for all } u, v, w \in \mathcal{H}(\mathbf{S}),$$

the metric space $(\mathcal{H}(\mathbf{S}), h)$ can now be formed. The completeness of the underlying metric space $(\mathbf{S}, d)$ is sufficient to show that every Cauchy sequence $\{u_k\}$ in $(\mathcal{H}(\mathbf{S}), h)$ converges to a point in $(\mathcal{H}(\mathbf{S}),h)$. Thus, $(\mathcal{H}(\mathbf{S}), h)$ is a *complete* metric space.

If S is metrizable, then it is $(\mathcal{H}(\mathbf{S}), h)$ wherein lurks that elusive beast, *semantic value*. For, consider the two propositions, $\rho_1 = (n_1, p_1)$, $\rho_2 = (n_2, p_2)$. Then the *nominative distance* $|n_2 - n_1|$ can be defined as $h(\overline{n_1}, \overline{n_2})$, where $\overline{n}$ denotes the closure of n. The *predicate distance* can be defined similarly. Finally, one might define:

$$|\rho_2 - \rho_1| = (|n_2 - n_1|^2 + |p_2 - p_1|^2)^{1/2} \qquad \text{Equation (1a)}$$

or alternatively one might use "city block" distance:

$$|\rho_2 - \rho_1| = |n_2 - n_1| + |p_2 - p_1| \qquad \text{Equation (1b)}$$

as a fair approximation of distance. Those skilled in the art will recognize that other metrics are also possible: for example:

$$(\Sigma(\rho_{2,i} - \rho_{1,i})^n)^{1/n} \qquad \text{Equation (1c)}$$

The reader may recognize $(\mathcal{H}(\mathbf{S}), h)$ as the *space of fractals*. Some compelling questions come immediately to mind. Might one be able to find *submonoids* of *contraction mappings* corresponding to *related* sets in $(\mathcal{H}(\mathbf{S}), h)$; related, for example, in the sense of convergence to the same collection of *attractors*? This could be a rich field to plow.

An Example Topology

Consider an actual topology on the set P of predicates. This is accomplished by exploiting the notion of hyponymy and meaning postulates.

Let P be the set of predicates, and let B be the set of all elements of 2^{2^P}, i.e., $\wp(\wp(P))$, that express hyponymy. B is a basis, if not of 2^P, i.e., $\wp(P)$, then at least of everything worth talking about: $S = \cup\ (b\colon b \in B)$. If $b_\alpha, b_\gamma \in B$, neither containing the other, have a non-empty intersection that is not already an explicit hyponym, extend the basis B with the meaning postulate $b_\alpha \cap b_\gamma$. For example, "dog" is contained in both "carnivore" and "mammal." So, even though the core lexicon may not include an entry equivalent to "carnivorous mammal," it is a worthy meaning postulate, and the lexicon can be extended to include the intersection. Thus, B is a basis for S.

Because hyponymy is based on nested subsets, there is a hint of partial ordering on S. A partial order would be a big step towards establishing a metric.

At this point, a concrete example of a (very restricted) lexicon is in order. FIG. 3 shows a set of concepts, including "thing" 305, "man" 310, "girl" 312, "adult human" 315, "kinetic energy" 320, and "local action" 325. "Thing" 305 is the maximal element of the set, as every other concept is a type of "thing." Some concepts, such as "man" 310 and "girl" 312 are "leaf concepts," in the sense that no other concept in the set is a type of "man" or "girl." Other concepts, such as "adult human" 315, "kinetic energy" 320, and "local action" 325 are "internal concepts," in the sense that they are types of other concepts (e.g., "local action" 325 is a type of "kinetic energy" 320) but there are other concepts that are types of these concepts (e.g., "man" 310 is a type of "adult human" 315).

FIG. 4 shows a directed set constructed from the concepts of FIG. 3. For each concept in the directed set, there is at least one chain extending from maximal element "thing" 305 to the concept. These chains are composed of directed links, such as links 405, 410, and 415, between pairs of concepts. In the directed set of FIG. 4, every chain from maximal element "thing" must pass through either "energy" 420 or "category" 425. Further, there can be more than one chain extending from maximal element "thing" 305 to any concept. For example, there are four chains extending from "thing" 305 to "adult human"

315: two go along link 410 extending out of "being" 435, and two go along link 415 extending out of "adult" 445.

Some observations about the nature of FIG. 4:

- First, the model is a *topological space.*
- Second, note that *the model is not a tree.* In fact, it is an example of a *directed set.* For example, concepts "being" 430 and "adult human" 315 are types of multiple concepts higher in the hierarchy. "Being" 430 is a type of "matter" 435 and a type of "behavior" 440; "adult human" 315 is a type of "adult" 445 and a type of "human" 450.
- Third, observe that the relationships expressed by the links are indeed relations of hyponymy.
- Fourth, note particularly – but without any loss of generality – that "man" 310 maps to both "energy" 420 and "category" 425 (via composite mappings) which in turn both map to "thing" 305; i.e., the (composite) relations are multiple valued and induce a partial ordering. These multiple mappings are natural to the meaning of things and critical to semantic characterization.
- Finally, note that "thing" 305 is *maximal*; indeed, "thing" 305 is the *greatest* element of *any* quantization of the lexical semantic field (subject to the premises of the model).

Metrizing S

FIGs. 5A-5G show eight different chains in the directed set that form a basis for the directed set. FIG. 5A shows chain 505, which extends to concept "man" 310 through concept "energy" 420. FIG. 5B shows chain 510 extending to concept "iguana." FIG. 5C shows another chain 515 extending to concept "man" 310 via a different path. FIGs. 5D-5G show other chains.

FIG. 13 shows a data structure for storing the directed set of FIG. 3, the chains of FIG. 4, and the basis chains of FIGs. 5A-5G. In FIG. 13, concepts array 1305 is used to store the concepts in the directed set. Concepts array 1305 stores pairs of elements. One element identifies concepts by name; the other element stores numerical identifiers 1306. For example, concept name 1307 stores the concept "dust," which is paired with numerical

Not shown in FIG. 13 is a data structure component for storing state vectors (discussed below). As state vectors are used in calculating the distances between pairs of concepts, if the directed set is static (i.e., concepts are not being added or removed and basis chains remain unchanged), the state vectors are not required after distances are calculated. Retaining the state vectors is useful, however, when the directed set is dynamic. A person skilled in the art will recognize how to add state vectors to the data structure of FIG. 13.

Although the data structure for concepts array 1305, maximal element 1310 chains array 1315, and basis chains array 1320 in FIG. 13 are shown as arrays, a person skilled in the art will recognize that other data structures are possible. For example, concepts array could store the concepts in a linked list, maximal element 1310 could use a pointer to point to the maximal element in concepts array 1305, chains array 1315 could use pointers to point to the elements in concepts array, and basis chains array 1320 could use pointers to point to chains in chains array 1315. Also, a person skilled in the art will recognize that the data in Euclidean distance matrix 1325A and angle subtended matrix 1325B can be stored using other data structures. For example, a symmetric matrix can be represented using only one half the space of a full matrix if only the entries below the main diagonal are preserved and the row index is always larger than the column index. Further space can be saved by computing the values of Euclidean distance matrix 1325A and angle subtended matrix 1325B "on the fly" as distances and angles are needed.

Returning to FIGs. 5A-5G, how are distances and angles subtended measured? The chains shown in FIGs. 5A-5G suggest that the relation between any node of the model and the maximal element "thing" 305 can be expressed as any one of a set of *composite* functions; one function for each chain from the minimal node μ to "thing" 305 (the n^{th} predecessor of μ along the chain):

$$f: \mu \Rightarrow \mathit{thing} = f_1 \circ f_2 \circ f_3 \circ \cdots \circ f_n$$

where the chain connects $n + 1$ concepts, and f_j: links the $(n - j)^{th}$ predecessor of μ with the $(n + 1 - j)^{th}$ predecessor of μ, $1 \leq j \leq n$. For example, with reference to FIG. 5A, chain 505 connects nine concepts. For chain 505, f_1 is link 505A, f_2 is link 505B, and so on through f_8 being link 505H.

Consider the set of all such functions for all minimal nodes. Choose a countable subset $\{f_k\}$ of functions from the set. For each f_k construct a function g_k: $S \Rightarrow \mathbf{I}^1$ as follows.

For $s \in S$, s is in relation (under hyponymy) to "thing" 305. Therefore, s is in relation to at least one predecessor of μ, the minimal element of the (unique) chain associated with f_k. Then there is a predecessor of smallest index (of μ), say the m^{th}, that is in relation to s. Define:

$$g_k(s) = (n - m) / n \qquad \text{Equation (2)}$$

This formula gives a measure of concreteness of a concept to a given chain associated with function f_k.

As an example of the definition of g_k, consider chain 505 of FIG. 5A, for which n is 8. Consider the concept "cat" 555. The smallest predecessor of "man" 310 that is in relation to "cat" 555 is "being" 430. Since "being" 430 is the fourth predecessor of "man" 310, m is 4, and g_k("cat" 555) = (8 - 4) / 8 = ½. "Iguana" 560 and "plant" 560 similarly have g_k values of ½. But the only predecessor of "man" 310 that is in relation to "adult" 445 is "thing" 305 (which is the eighth predecessor of "man" 310), so m is 8, and g_k("adult" 445) = 0.

Finally, define the vector valued function $\varphi: S \Rightarrow \mathbb{R}^k$ relative to the indexed set of scalar functions $\{g_1, g_2, g_3, \ldots, g_k\}$ (where scalar functions $\{g_1, g_2, g_3, \ldots, g_k\}$ are defined according to Equation (2)) as follows:

$$\varphi(s) = \langle g_1(s), g_2(s), g_3(s), \ldots, g_k(s)\rangle \qquad \text{Equation (3)}$$

This state vector $\varphi(s)$ maps a concept s in the directed set to a point in k-space ($\mathbb{R}^k$). One can measure distances between the points (the state vectors) in k-space. These distances provide measures of the closeness of concepts within the directed set. The means by which distance can be measured include distance functions, such as Equations (1a), (1b), or (1c). Further, trigonometry dictates that the distance between two vectors is related to the angle subtended between the two vectors, so means that measure the angle between the state vectors also approximates the distance between the state vectors. Finally, since only the direction (and not the magnitude) of the state vectors is important, the state vectors can be normalized to the unit sphere. If the state vectors are normalized, then the angle between two state vectors is no longer an approximation of the distance between the two state vectors, but rather is an exact measure.

The functions g_k are analogous to step functions, and in the limit (of refinements of the topology) the functions are continuous. Continuous functions preserve local topology; i.e., "close things" in S map to "close things" in $\mathbb{R}^k$, and "far things" in S tend to map to "far things" in $\mathbb{R}^k$.

Example Results

The following example results show state vectors $\varphi(s)$ using chain 505 as function g_1, chain 510 as function g_2, and so on through chain 540 as function g_8.

$\varphi(\text{"boy"}) \Rightarrow \langle 3/4,\ 5/7,\ 4/5,\ 3/4,\ 7/9,\ 5/6,\ 1,\ 6/7 \rangle$

$\varphi(\text{"dust"}) \Rightarrow \langle 3/8,\ 3/7,\ 3/10,\ 1,\ 1/9,\ 0,\ 0,\ 0 \rangle$

$\varphi(\text{"iguana"}) \Rightarrow \langle 1/2,\ 1,\ 1/2,\ 3/4,\ 5/9,\ 0,\ 0,\ 0 \rangle$

$\varphi(\text{"woman"}) \Rightarrow \langle 7/8,\ 5/7,\ 9/10,\ 3/4,\ 8/9,\ 2/3,\ 5/7,\ 5/7 \rangle$

$\varphi(\text{"man"}) \Rightarrow \langle 1,\ 5/7,\ 1,\ 3/4,\ 1,\ 1,\ 5/7,\ 5/7 \rangle$

Using these state vectors, the distances between concepts and the angles subtended between the state vectors are as follows:

Pairs of Concepts	Distance (Euclidean)	Angle Subtended
"boy" and "dust"	~1.85	~52°
"boy" and "iguana"	~1.65	~46°
"boy" and "woman"	~0.41	~10°
"dust" and "iguana"	~0.80	~30°
"dust" and "woman"	~1.68	~48°
"iguana" and "woman"	~1.40	~39°
"man" and "woman"	~0.39	~07°

From these results, the following comparisons can be seen:

- "boy" is closer to "iguana" than to "dust."
- "boy" is closer to "iguana" than "woman" is to "dust."
- "boy" is much closer to "woman" than to "iguana" or "dust."
- "dust" is further from "iguana" than "boy" to "woman" or "man" to "woman."
- "woman" is closer to "iguana" than to "dust."

- “woman” is closer to “iguana” than “boy” is to “dust.”
- “man” is closer to “woman” than “boy” is to “woman.”

All other tests done to date yield similar results. The technique works consistently well.

How It (Really) Works

As described above, construction of the φ transform is (very nearly) an algorithm. In effect, this describes a *recipe* for metrizing a lexicon – or for that matter, metrizing anything that can be modeled as a directed set – but does not address the issue of *why* it works. In other words, *what's really going on here*? To answer this question, one must look to the underlying mathematical principles.

First of all, what is the nature of **S**? Earlier, it was suggested that a propositional model of the lexicon has found favor with many linguists. For example, the lexical element “automobile” might be modeled as:

{automobile: *is a* machine,
is a vehicle,
has engine,
has brakes,
...
}

In principle, there might be infinitely many such properties, though practically speaking one might restrict the cardinality to $\aleph_0$ (countably infinite) in order to ensure that the properties are addressable. If one were disposed to do so, one might require that there be only finitely many properties associated with a lexical element. However, there is no compelling reason to require finiteness.

At any rate, one can see that “automobile” is simply an element of the power set of **P**, the set of all propositions; i.e., it is an element of the set of all subsets of **P**. The power set is denoted as $\wp(\mathbf{P})$. Note that the first two properties of the “automobile” example express “*is a*” relationships. By “*is a*” is meant entailment. *Entailment* means that, were one to intersect the properties of every element of $\wp(\mathbf{P})$ that is called, for example, “machine,” then the

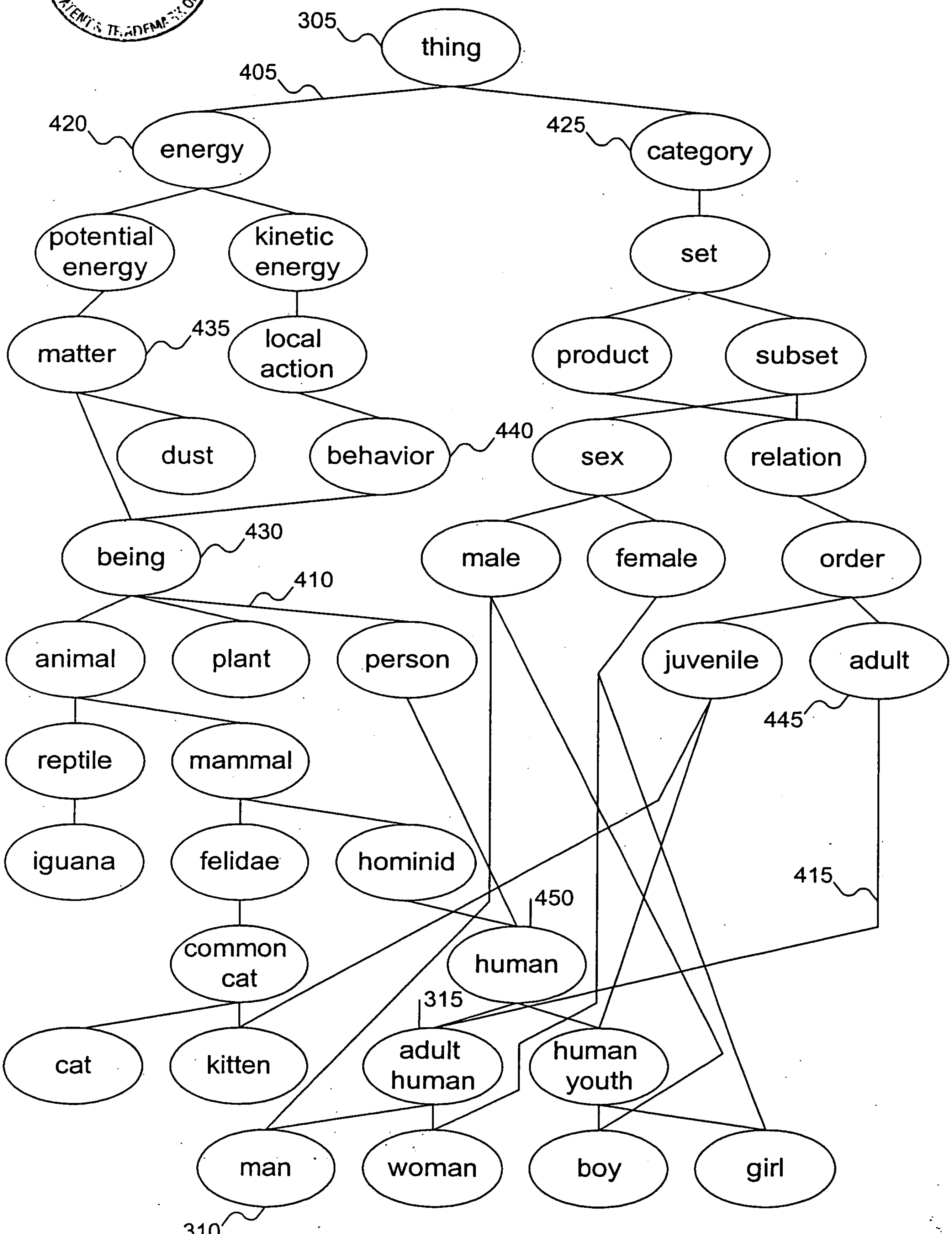

305
thing
405
420
energy
425
category
potential energy
kinetic energy
set
matter
435
local action
product
subset
dust
behavior
440
sex
relation
being
430
male
female
order
410
animal
plant
person
juvenile
adult
445
reptile
mammal
iguana
felidae
hominid
415
450
common cat
human
315
cat
kitten
adult human
human youth
man
woman
boy
girl
310

FIG. 4

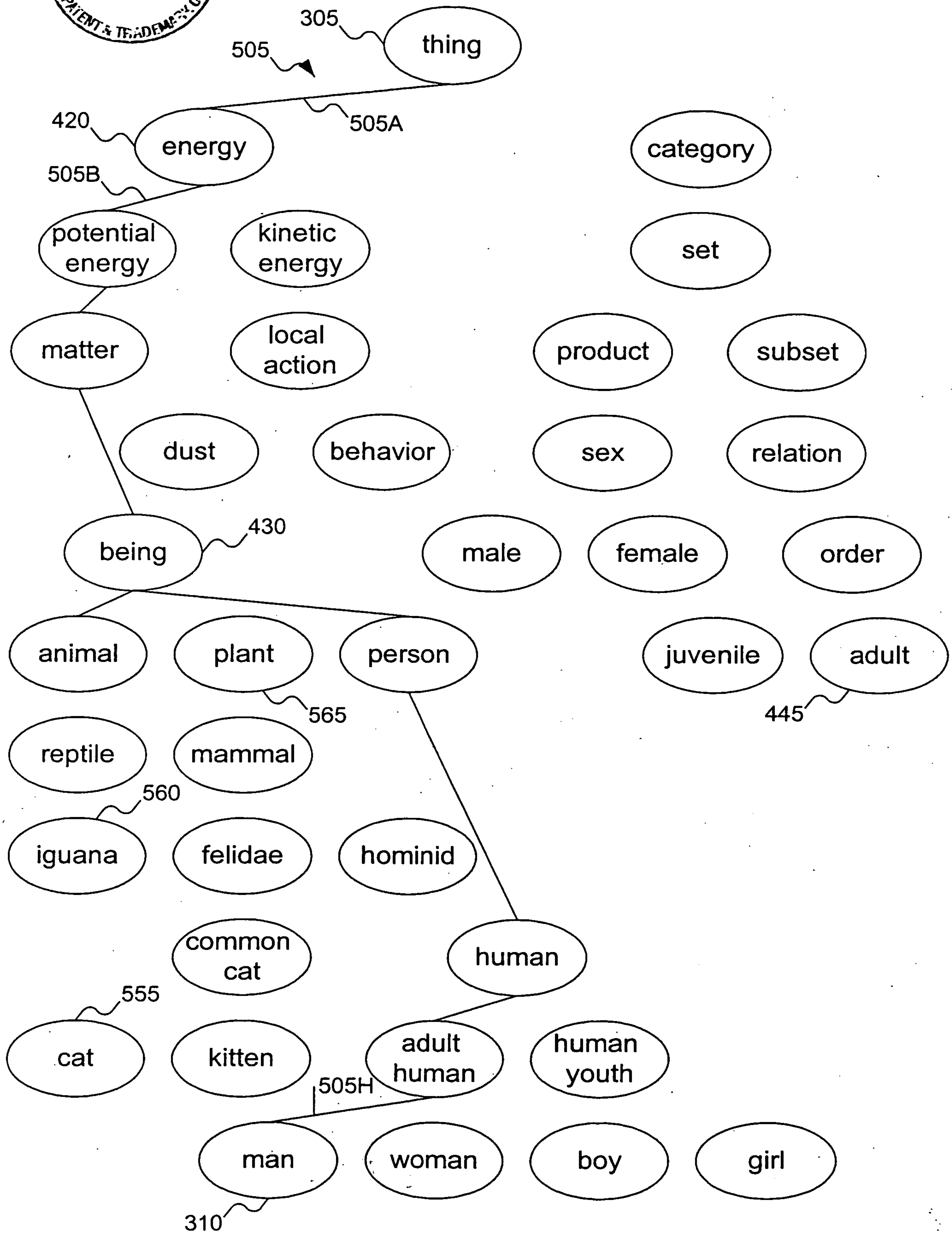
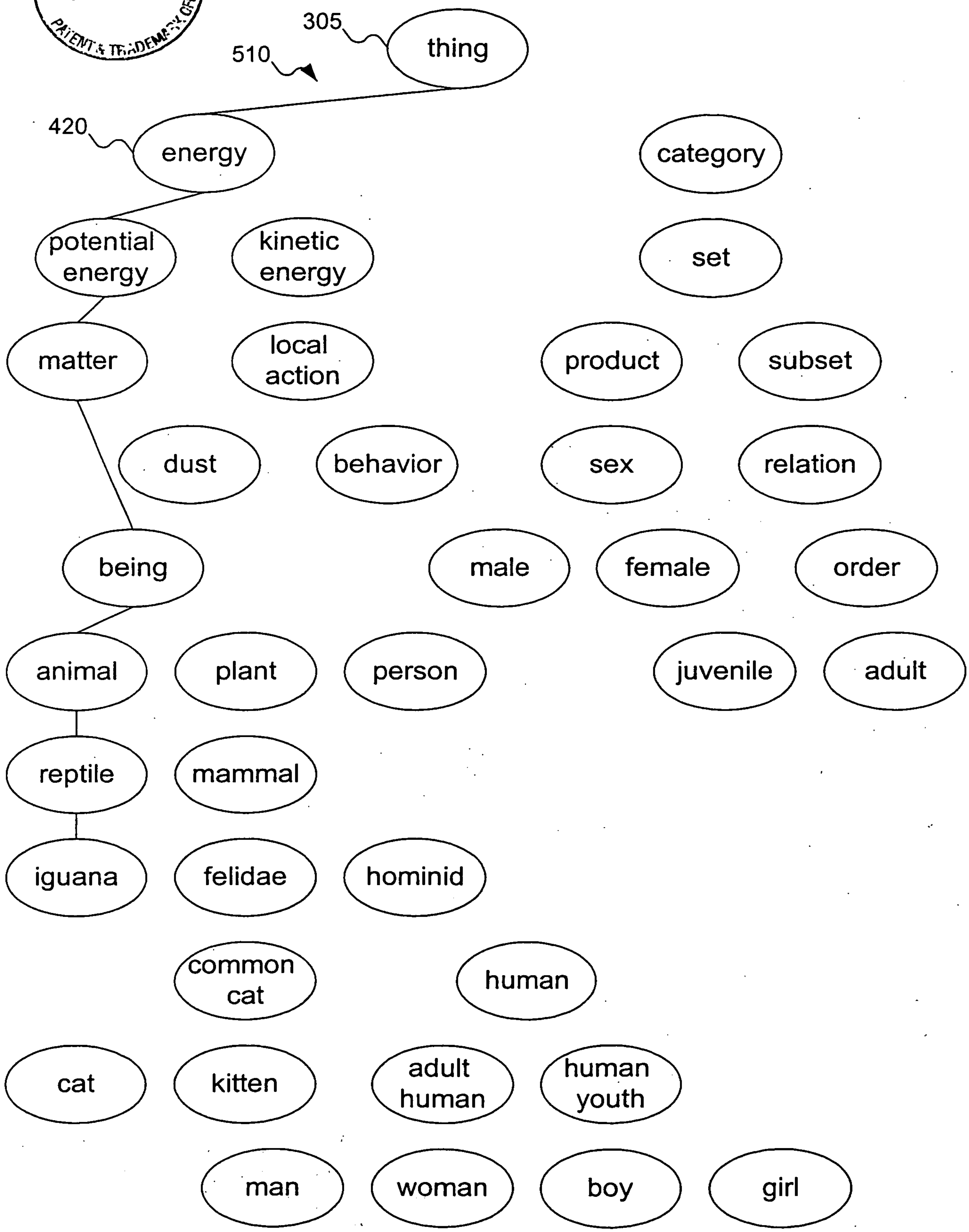

305
thing
505
505A
420
energy
category
505B
potential energy
kinetic energy
set
matter
local action
product
subset
dust
behavior
sex
relation
being
430
male
female
order
animal
plant
person
juvenile
adult
565
445
reptile
mammal
560
iguana
felidae
hominid
common cat
human
555
cat
kitten
adult human
human youth
505H
man
woman
boy
girl
310

FIG. 5A



305
510
thing
420
energy
category
potential energy
kinetic energy
set
matter
local action
product
subset
dust
behavior
sex
relation
being
male
female
order
animal
plant
person
juvenile
adult
reptile
mammal
iguana
felidae
hominid
common cat
human
cat
kitten
adult human
human youth
man
woman
boy
girl

FIG. 5B

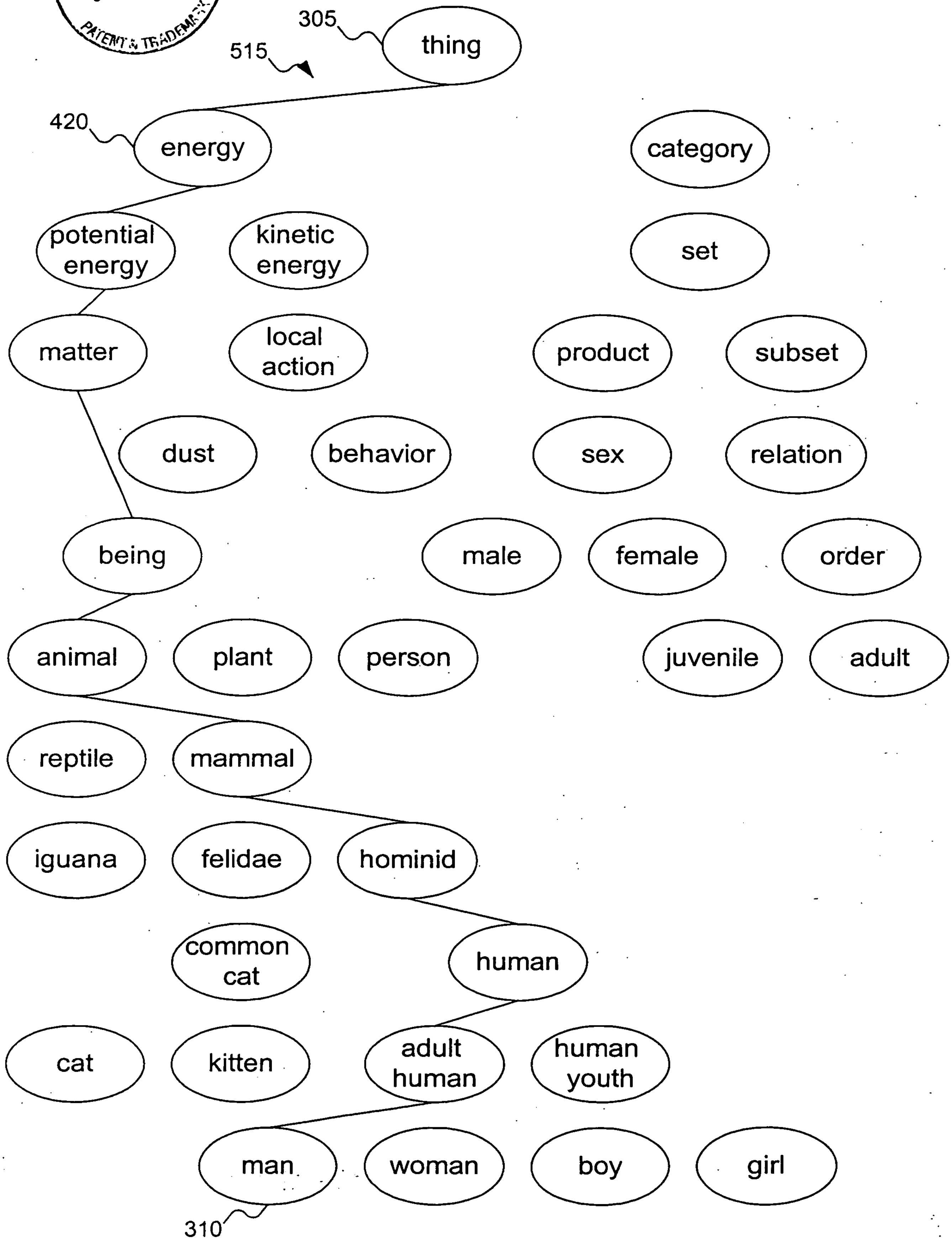

305
thing
515
420
energy
category
potential energy
kinetic energy
set
matter
local action
product
subset
dust
behavior
sex
relation
being
male
female
order
animal
plant
person
juvenile
adult
reptile
mammal
iguana
felidae
hominid
common cat
human
cat
kitten
adult human
human youth
man
woman
boy
girl
310

FIG. 5C

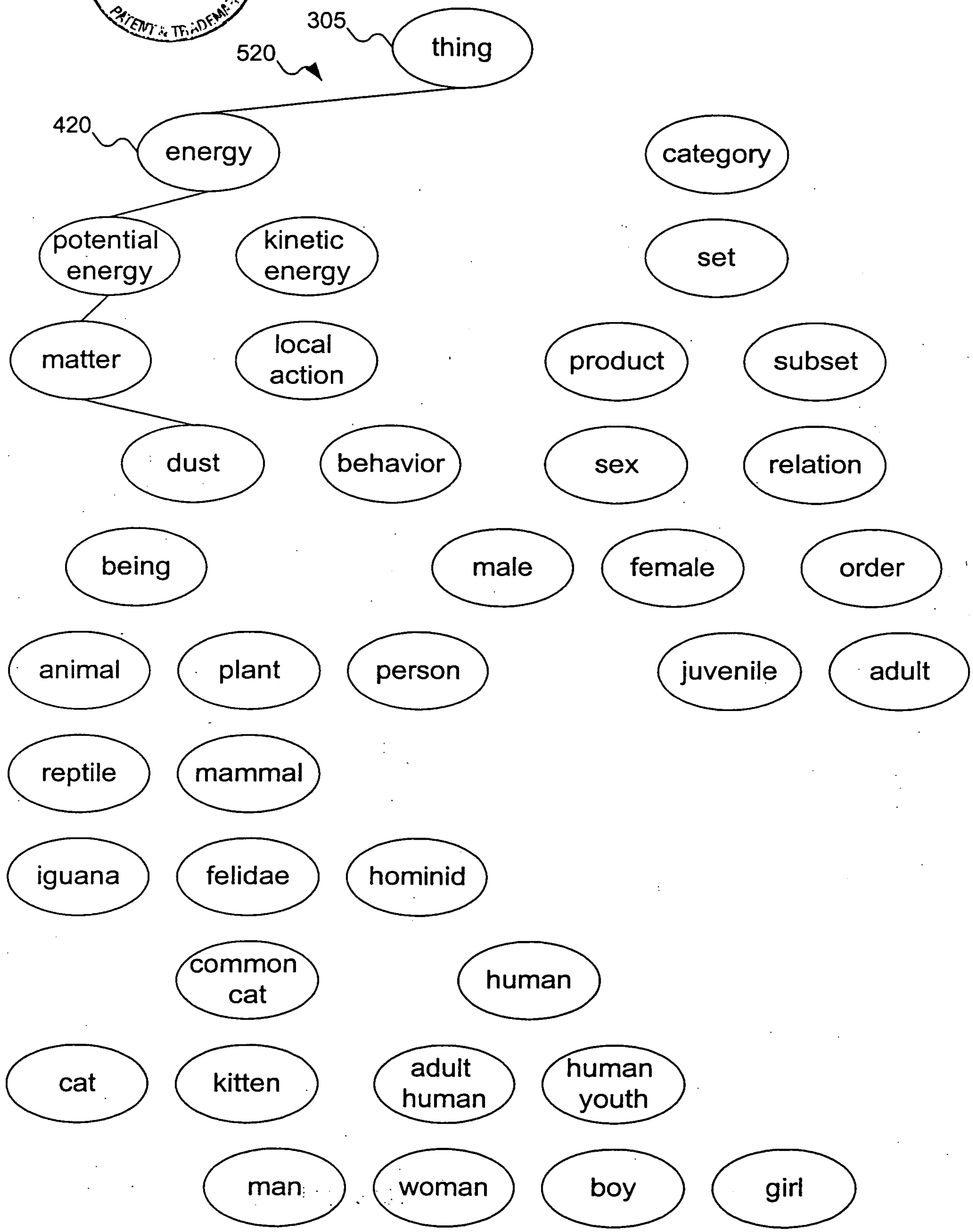

305
thing
520
420
energy
category
potential energy
kinetic energy
set
matter
local action
product
subset
dust
behavior
sex
relation
being
male
female
order
animal
plant
person
juvenile
adult
reptile
mammal
iguana
felidae
hominid
common cat
human
cat
kitten
adult human
human youth
man
woman
boy
girl

FIG. 5D

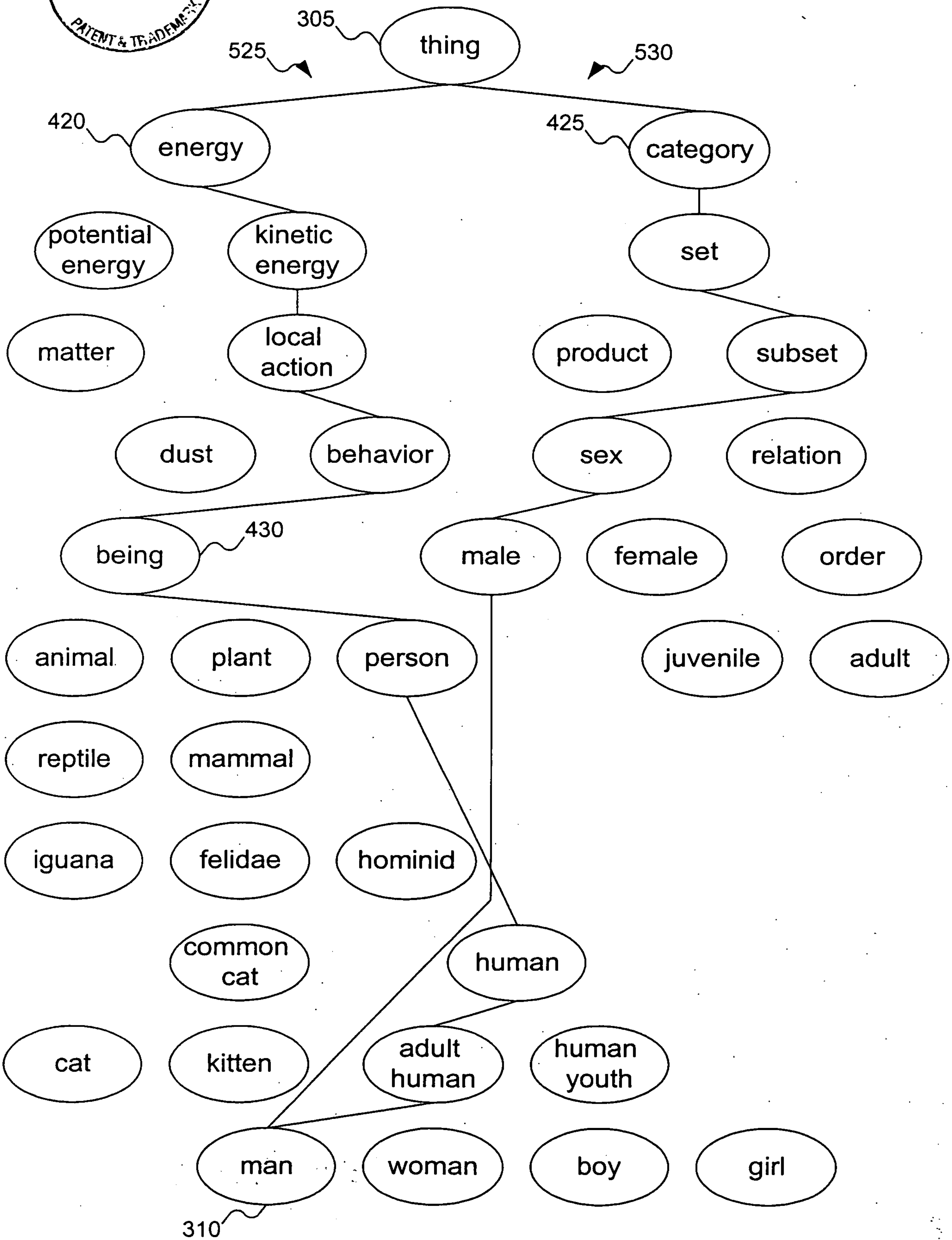

305
thing
525
530
420
energy
425
category
potential energy
kinetic energy
set
matter
local action
product
subset
dust
behavior
sex
relation
being
430
male
female
order
animal
plant
person
juvenile
adult
reptile
mammal
iguana
felidae
hominid
common cat
human
cat
kitten
adult human
human youth
man
woman
boy
girl
310

FIG. 5E

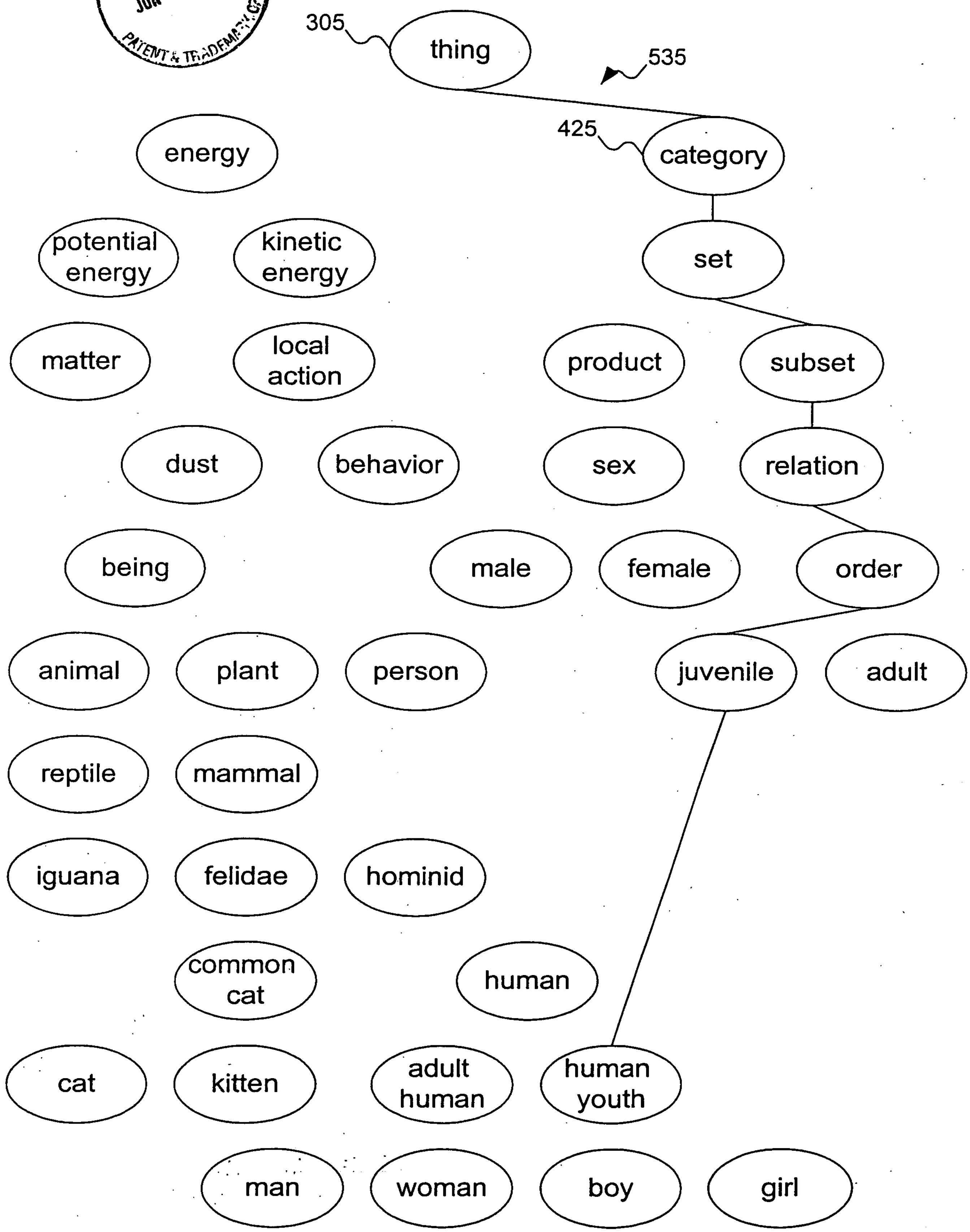

305
thing
535
425
category
energy
potential energy
kinetic energy
set
matter
local action
product
subset
dust
behavior
sex
relation
being
male
female
order
animal
plant
person
juvenile
adult
reptile
mammal
iguana
felidae
hominid
common cat
human
cat
kitten
adult human
human youth
man
woman
boy
girl

FIG. 5F

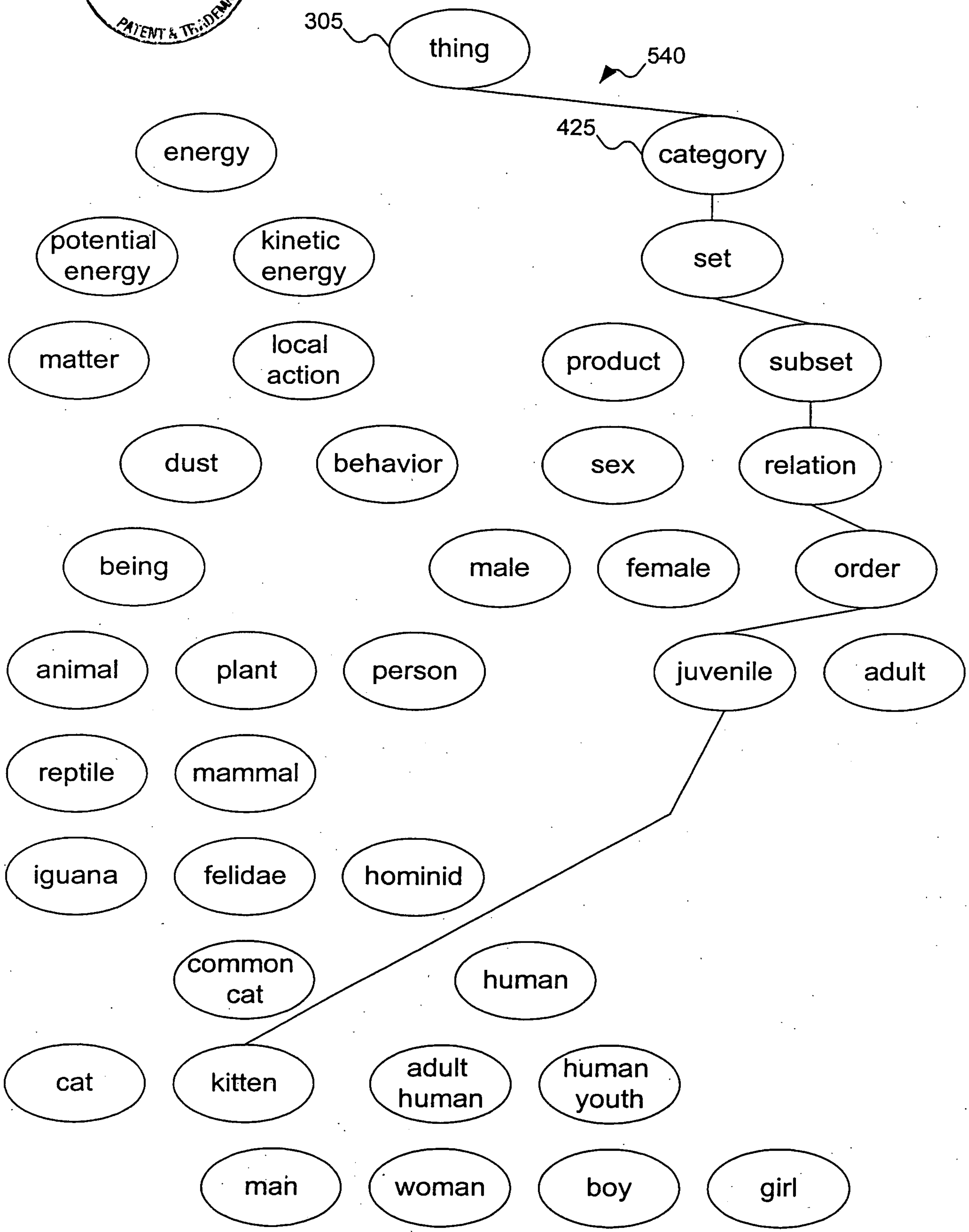

305
thing
540
425
category
energy
potential energy
kinetic energy
set
matter
local action
product
subset
dust
behavior
sex
relation
being
male
female
order
animal
plant
person
juvenile
adult
reptile
mammal
iguana
felidae
hominid
common cat
human
cat
kitten
adult human
human youth
man
woman
boy
girl

FIG. 5G